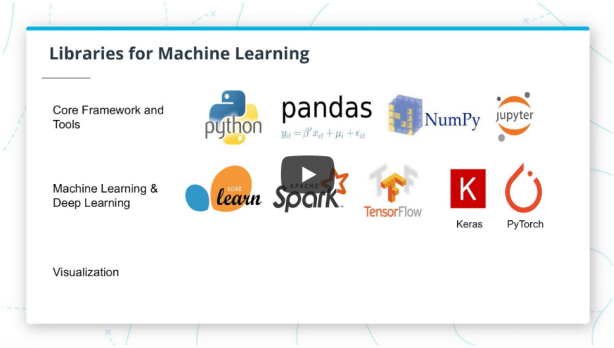
Libraries for Machine Learning



For your reference, here are all the libraries we went over in the video. This is a lot of info; you should not feel like you need to be deeply knowledgable about every detail of these libraries. Rather, we suggest that you become familiar with what each library is *for*, in general terms. For example, if you hear someone talking about *matplotlib*, it would be good for you to recognize that this is a popular library for data visualization. Or if you see a reference to *TensorFlow*, it would be good to recognize this as a popular machine learning library.

Core Framework and Tools

- Python is a very popular high-level programming language that is great for data science. Its ease of use and wide support within popular machine learning platforms, coupled with a large catalog of ML libraries, has made it a leader in this space.
- Pandas is an open-source Python library designed for analyzing and manipulating data. It is particularly good for working with tabular data and time-series data.
- NumPy, like Pandas, is a Python library. NumPy provides support for large, multi-dimensional arrays of data, and has many high-level mathematical functions that can be used to perform operations on these arrays.

Machine Learning and Deep Learning

- Scikit-Learn is a Python library designed specifically for machine learning. It is designed to be integrated with other scientific and data-analysis libraries, such as NumPy, SciPy, and matplotlib (described below).
- Apache Spark is an open-source analytics engine that is designed for cluster-computing and that is often used for large-scale data processing and big data.
- TensorFlow is a free, open-source software library for machine learning built by Google Brain.
- Keras is a Python deep-learning library. It provide an Application Programming Interface (API) that can be used to interface with other libraries, such as TensorFlow, in order to program neural networks. Keras is designed for rapid development and experimentation.
- PyTorch is an open source library for machine learning, developed in large part by Facebook's AI Research lab. It is known for being comparatively easy to use, especially for developers already familiar with Python and a Pythonic code style.

Data Visualization

- Plotly is not itself a library, but rather a company that provides a number of different front-end tools for machine learning and data science—including an open source graphing library for Python.
- Matplotlib is a Python library designed for plotting 2D visualizations. It can be used to produce graphs and other figures that are high quality and usable in professional publications. You'll see that the Matplotlib library is used by a number of other libraries and tools, such as SciKit Learn (above) and Seaborn (below). You can easily import Matplotlib for use in a Python script or to create visualizations within a Jupyter Notebook.
- Seaborn is a Python library designed specifically for data visualization. It is based on matplotlib, but provides a more high-level interface and has additional features for making visualizations more attractive and informative.
- Bokeh is an interactive data visualization library. In contrast to a library like matplotlib that generates a static image as its output, Bokeh generates visualizations in HTML and JavaScript. This allows for web-based visualizations that can have interactive features.

QUIZ QUESTION

Below are some of the libraries we just went over. See if you can match each library with its main focus.

Submit to check your answer choices!

LIBRARY	WHAT IS IT FOR?
TensorFlow	Machine learning
Matplotlib	Data visualization
Pandas	Analyzing/manipulating data
PyTorch	Machine learning
Bokeh	Data visualization

SUBMIT

NEXT